

Population Estimates for Connecticut

Prepared by Karyn Backus and Lloyd Mueller
CT Department of Public Health



Introduction

If you asked three people to find the population of your town for a given year, you would probably expect that each person would report back the same number. After all, there is only one actual population count for your town for a given year, you might think. Instead, each person reported back with a different number!

Welcome to Population Estimates. With this document, I strive to educate you on how population estimates are derived. This complexity can result in different numbers for the same point in time. The complexity also means that one dataset may be a better fit for your project than another. Understanding the differences between the various population estimates will allow you to choose the population dataset that is most appropriate for your needs.

Before moving on, it is important to convey that **population numbers are ALWAYS estimates**. Since it is impossible to count each and every person in the United States at a single point in time, the population data available to us are always an *estimate* of the true population. By its very nature, estimation is an inexact science and the outcome of the estimate is dependent on the methods that were applied to create the estimate. No one knows the actual true count of the population at any given time, but several estimation methods have been developed in order to provide to us a reasonable approximation of the actual count. These multiple methodologies lead to multiple population estimates for the same population which leads to three people providing a different count for the same town and year. None of these three persons are wrong per se, but how you will be using the population estimates may dictate that one estimate is more appropriate than another.

In Public Health, populations as a whole are of limited interest. Drilling down to identify population subsets based on age, sex, race and/or Hispanic ethnicity are often a significant component of our evaluations. When choosing a dataset, the methodology used is important to consider, yet the level of demographic breakdown can be the defining factor in which estimate is ultimately chosen. These interests can be competing as one dataset may be more accurate but fails to provide the necessary level of detail (e.g., age*sex breakdown). The necessity of an age*sex breakdown may lead you to use a less accurate estimate in order to meet the demographic demands of your project.

In the next few pages, I briefly review the various population estimates datasets that are utilized by the CT Department of Public Health. All of the population estimates originate from the United States Census Bureau (USCB). The decennial census is the foundation for all of the estimation methods. After the decennial census, the USCB

provides annual estimates of the population. The National Center for Health Statistics (NCHS) provides modified versions of the USCB's annual estimates. Once another decennial census is completed, the USCB publishes revisions to the annual estimates between the two decennial censuses. At the same time that the USCB is providing annual estimates, the USCB is also conducting the American Community Survey which provides annual, 3-year, and 5-year estimates of the population with limited demographic detail. This results in a variety of population estimates - each of which is slightly different.

The amount of population data available can be overwhelming. Once you narrow your scope by identifying any required demographic subgroups, by identifying the years of interest, and/or by identifying the level of geography, the amount of population estimates available is greatly reduced. The following pages discuss each of the available population estimate sources along with their level of demographic and geographic breakdown.

Decennial Census Counts

The estimation process that is closest to providing a true count of our Nation is the *decennial census*. The US Census Bureau (USCB) takes a census of the population of the United States every 10 years which is known as the decennial census. A major source of population data for CT, the decennial census *attempts* to provide a *true count of the population rather than an estimate*. By conducting an actual census, this method is the best estimation of a true count.

Although the census is described as a count, the population numbers provided to us are still estimates. The USCB is unable to truly count each person accurately on April 1st. To account for the failure of some persons to be counted, adjustments are made to the true counts to generate population numbers that better represent a full count. In the end, the decennial census is an estimate based off of a true count.

Strengths:

- Considered the best representation of the actual population.
- Provides estimates for every level of geography supported by the USCB.
 - Nation, State, County, MCD (Town), Tract, Block Group, Block and other political units
- Provides demographic estimates that can be cross-tabulated to access specific demographic sub-groups at every level of geography.
 - Counts by age, sex, race and ethnicity are available at the town level.
- [American FactFinder](#) provides easy access to data.

Weaknesses:

- Estimates are for April 1st
 - All other population estimates use July 1st making the decennial census inconsistent
 - The population change between April 1 and July 1 may or may not lead to underrepresentation of the population that is large enough to affect analyses
- Data is published *as reported* by the respondents
 - Corrections and reallocations are not applied to the decennial census datasets
 - Extra Race categories: Some Other Race & Two or More Races
 - In addition to the standard race groups (White, Black, American Indian/Alaskan Native, Asian, Hawaiian and Pacific Islander), the Census also collects data for Multiple Races and Some Other Race.
 - These Multiple Race and Some Other Race categories can make it difficult to use decennial Census data as denominators when analyzing numerator data that does not contain these extra categories.
 - The tendency for Hispanic persons to self-identify as Some Other Race means that Hispanics are significantly under-counted in the standard race groups.

Town ASRH:

- **The USCB publishes town-level demographic data for April 1, 2010; however, this dataset has several major limitations that undermine its utility (see weaknesses above). Use with caution.**

Census 2010 datasets available @ <http://www.ct.gov/dph/populationdata>:

[Connecticut Profile of General Population and Housing Characteristics, 2010](#)

This profile provides population figures from April 1, 2010 by state, county and town with some demographic breakdowns.

[Connecticut Town Population by Age Sex Race and Hispanic Ethnicity, 2010](#)

This dataset contains population figures from April 1, 2010 by town, single year of age, sex, race group, and Hispanic status. (xls, 9,804KB)
(Source: [Census 2010 Summary File 1](#))

Annual Post-censal Estimates

The populations of Connecticut at the state, county and town level are estimated at the calendar year midpoint (July 1) each year and are known as *annual* or *mid-year* estimates. Annual estimates are also known as *post-censal* estimates because the USCB uses a variety of population change factors (births, deaths, Federal tax returns, Medicare enrollment, and immigration) to update the decennial census base counts from April 1, 2010. Once the inputs are updated, the estimates for each year back to the decennial census are revised and this new series of population estimates are published. The latest year in each series denotes the vintage of the dataset.

Although the April 1, 2010 census is an actual count of the population on that day, a mid-year estimate (July 1) better represents the population count for an entire year. Thus, annual *post-censal* estimates are created to serve as the best representation of a single year's population and to provide a consistent reference point across years. The Population Estimates Program (PEP) of the USCB publishes the annual, post-censal datasets about a year following the estimate date (e.g., July 1, 2012 data will be published around July 1, 2013).

Connecticut General Statutes mandates that CT DPH provide the Office of Policy and Management the state, county, and town population totals. These totals are used throughout the State to distribute per capita funds. Once the USCB releases their annual estimate, our DPH reviews the USCB's annual estimates and makes any necessary adjustments. Our DPH then releases the official population estimates for Connecticut at the state, county, and town levels by October of the year following the estimate date.

Included with each release of data is a discussion of the methodology used to calculate the estimates with special attention to changes in the USCB methodology and to any DPH adjustments. Usually the official CT estimates are identical to the USCB's publication; however, occasionally, there has been cause for DPH to adjust the USCB's figures resulting in CT official estimates that are not the same as the USCB's. Due to the potential modifications, it is important to use the official CT estimates whenever your data should be consistent with State of Connecticut publications.

Using the official CT estimates is always a reasonable option when trying to choose an appropriate population estimate dataset. **First**, these are the official estimates so it is easy to rationalize why these were used for your rates. **Second**, the official CT estimates are static. Unlike other series which are updated each year, the official CT estimates come from the first vintage published for that year. They are not updated or revised with subsequent vintages, which keeps them static. **Third**, there may be times when consistency with prior estimates is a key concern. For instance, if users will be comparing new disease rates with rates based on official CT estimates, it makes sense to use the original CT estimates for the new disease rates as well. Maintaining the same denominator series minimizes confusion. The gains in accuracy from a newer series of estimates may be lost if the revised rates create confusion among users of those statistics. Even small changes in the value of a rate can create significant confusion when specific values have been previously quoted (e.g., to the news media, legislators, the commissioner), but then those values are no longer available. Both the

costs and benefits should be considered before modifying the population denominator series used to calculate rates.

Strengths:

- Based off of the decennial census.
- Provides estimates for July 1st.
- The Some Other Race category has been reallocated so these estimates better represent Hispanic numbers.
- Corrections to inputs (births, deaths, migration, etc.) are applied

Weaknesses:

- Does not provide any estimates for geographies lower than MCD (Town), such as tracts.
- Demographic breakdowns are not available at the MCD (Town) level.
- Demographic breakdowns that are available for County-level and up are not as complete as the decennial census.
- The Two or More Race category is retained.
- Inputs (births, deaths, migration) are often delayed limiting true-to-date estimates
- A new vintage is published each year
 - Revision of the previous estimates each time a new year is added to the series creates multiple datasets for any given year. These multiple datasets will have different estimates for the same year which may cause confusion.
 - Methodology may change over time. Changes in the estimation process may obfuscate the actual trend in population change when years from different vintages are compared.
 - Data citations must specify the vintage of the dataset.

Town ASRH:

- The USCB publishes town-level population totals for July 1st annually; however, no demographic data is provided.
- CT DPH developed demographic data at the town-level for July 1, 2000 and July 1, 2010 (discussed in subsequent sections of this document).

Annual, post-censal datasets available @ <http://www.ct.gov/dph/populationdata>:

[Town & County: Population \(without demographic breakdowns\)](#)

Special notes about data availability and population changes:

Changes in Estimates Methodology: The Vintage 2009 population estimates reflect four major improvements in the estimates methodology. At the national level, the result of the various methodological changes is an upward shift of the Vintage 2009 post-censal population estimates when compared to those from the Vintage 2008 series. At the State and county level, some race and age groups experienced substantial changes (comparison of July 1, 2008 estimates from the Vintage 2008 series with the July 1, 2008 and July 1, 2009 estimates from the Vintage 2009 series). The Vintage 2008 post-censal population estimates reflect three major improvements in the estimates methodology. These methodology changes go beyond the extensive methodology changes implemented for the Vintage 2007 estimates. The net impact of the various methodological changes is a downward shift of the Vintage 2008 post-censal population estimates when compared to those from the Vintage 2007 series; the Vintage 2007 post-censal estimates also reflected a downward shift when compared to the 2006 estimates.

Annual Bridged-ASRH Post-censal Estimates

The National Center for Health Statistics (NCHS) has developed a method to produce annual estimates at the state and county levels by age, sex, race and Hispanic ethnicity (ASRH) using four single-race categories instead of the 31 single and multiple-race categories published by the U.S. Census in 2000 and 2010.

The NCHS bridged estimates are created directly from the USCB's annual post-censal estimates. Through a process known as 'bridging', the multiple race groups are partially reallocated into single race groups to produce annual post-censal population estimates using four mutually-exclusive race categories (White, Black, American Indian/Alaskan Native, and Asian/Pacific Islander).

NCHS' bridged population estimates are particularly useful since many health data systems still collect information using single-race categories (White, Black, American Indian/Alaskan Native, and Asian/Pacific Islander). By providing these bridged race categories, DPH can calculate race-specific rates to monitor many health indicators that would have been difficult to evaluate using the unbridged, multiple race population estimates.

NCHS' ASRH data contains estimates annually as of July 1 by state/county, single year of age, sex, race (White, Black/African American, American Indian/Alaska Native, Asian/Pacific Islander) and ethnicity (Hispanic/Latino, not Hispanic/Latino).

Strengths:

- Same strengths as the annual post-censal estimates because the data inputs are identical:
 - Based off of the decennial census.
 - Provides estimates for July 1st.
 - The Some Other Race category has been reallocated so these estimates better represent Hispanic numbers.
 - Corrections to inputs (births, deaths, migration, etc.) are applied.
- Conforms to the 1977 OMB Standards which may be more aligned with DPH data collection categories for race.
 - Reallocates multiple race groups into single race groups
 - Creates denominator categories that are consistent with numerator categories
 - Fewer demographic (ASRH) combinations with a cell size of zero.

Weaknesses:

- Same weaknesses as the annual post-censal estimates because the data inputs are identical:
 - Does not provide any estimates for geographies lower than MCD (Town), such as tracts.
 - Demographic breakdowns are not available at the MCD (Town) level.
 - Demographic breakdowns that are available for County-level and up are not as complete as the decennial census.
 - Inputs (births, deaths, migration) are often delayed limiting true-to-date estimates
 - A new vintage is published each year
 - Revision of the previous estimates each time a new year is added to the series creates multiple datasets for any given year. These multiple datasets will have different estimates for the same year which may cause confusion.
 - Methodology may change over time. Changes in the estimation process may obfuscate the actual trend in population change when years from different vintages are compared.
 - Data citations must specify the vintage of the dataset.

- Collapsing the Asian race group with the Native Hawaiian and Other Pacific Islander race group may be a weakness for some users.
- County-level ASRH is not available for vintage 2000 or vintage 2001.

Town ASRH:

- **NCHS does not publish ASRH data at the town-level because the USCB does not publish data at the town-level.**

Annual, bridged-ASRH post-censal datasets available @
<http://www.ct.gov/dph/populationdata>:

[State & County: Population by Age, Sex, Race and Hispanic Ethnicity \(ASRH\)](#)

Special notes about data availability and population changes:

County level ASRH estimates that are consistent with the original state level figures are not available for 2000 and 2001. NCHS did not publish July 1, 2000 estimates or July 1, 2001 estimates at the county level when the original state level estimates were produced. County level estimates for 2000 and 2001 were produced at a later date, with the release of the 2002 estimates. However, these "vintage 2002" data include modified county level 2000 and 2001 estimates that differ from the official CT State Population estimates.

Changes in Estimates Methodology: The Vintage 2009 population estimates reflect four major improvements in the estimates methodology. At the national level, the result of the various methodological changes is an upward shift of the Vintage 2009 post-censal population estimates when compared to those from the Vintage 2008 series. At the State and county level, some race and age groups experienced substantial changes (comparison of July 1, 2008 estimates from the Vintage 2008 series with the July 1, 2008 and July 1, 2009 estimates from the Vintage 2009 series). The Vintage 2008 post-censal population estimates reflect three major improvements in the estimates methodology. These methodology changes go beyond the extensive methodology changes implemented for the Vintage 2007 estimates. The net impact of the various methodological changes is a downward shift of the Vintage 2008 post-censal population estimates when compared to those from the Vintage 2007 series; the Vintage 2007 post-censal estimates also reflected a downward shift when compared to the 2006 estimates.

July 1, 2000 & July 1, 2010 Town-level Bridged-ASRH Post-censal Estimates

Population estimates by age, sex, race and Hispanic ethnicity (ASRH) are invaluable for the estimation of rates of health and illness in Connecticut's towns and counties. Unfortunately, the annual post-censal town population estimates do not include the ASRH components *and* the population change between April 1 and July 1 is substantial enough to prohibit the direct substitution of demographic counts. To address this, our DPH developed a town-level ASRH dataset for July 1, 2000.

This town-level dataset provides population estimates that are consistent with the official mid-year state and town estimates of the population for July 1, 2000 that provides a consistent reference point with the annual estimates for 1991-1999 and 2001-2009. These estimates also allow rates of health and illness to be calculated at the town level using the mid-year 2000 population rather than the lower population estimate from the Census in April.

This dataset was created by taking the demographic distribution of age, sex, race, and ethnicity by town for April 1, 2000 and adjusting it to fit both the town population totals for July 1, 2000 and the county demographic totals by age, sex, bridged race/ethnicity. The result is a town-level bridged ASRH dataset where the totals are consistent with other existing July 1, 2000 datasets.

It is important to note that the town-level ASRH uses a collapsed race*ethnicity structure. The race/ethnicity groups remain mutually-exclusive, but they are now conjoined by Hispanic status. The four race groups are provided for non-Hispanic persons only and the Hispanic ethnicity is provided for Hispanic persons of any race. This structure results in 5 mutually-exclusive race/ethnicity categories: White-NH, Black-NH, American Indian-NH, Asian/Pacific Islander-NH, or Hispanic/Latino-any race.

The town-level ASRH July 1, 2000 file contains estimates of the resident population of Connecticut as of July 1, 2000 by town, five-year age group, sex, mutually exclusive race/ethnicity categories (White-NH, Black/African American-NH, American Indian/Alaska Native-NH, Asian/Pacific Islander-NH, or Hispanic/Latino-any race).

After the 2010 decennial Census, DPH created a 2010 town-level ASRH dataset in the same manner.

Strengths:

- The only dataset that provides a single-year, town-level estimate with demographic cross-tabulations for a July 1st time point.
- Same strengths as the annual post-censal estimates because the data inputs are identical:
 - Based off of the decennial census.
 - Provides estimates for July 1st.
 - The Some Other Race category has been reallocated so these estimates better represent Hispanic numbers.
 - Corrections to inputs (births, deaths, migration, etc.) are applied.
- Conforms to the 1977 OMB Standards which may be more aligned with DPH data collection categories for race.
 - Reallocates multiple race groups into single race groups
 - Creates denominator categories that are consistent with numerator categories
 - Collapses the Hispanic ethnicity into a single race*ethnicity category

- Substantially reduces the number of Hispanic age*sex*race combinations with a cell size of zero.
- More consistent with numerator data which often collapses Hispanic persons into a single race*ethnicity group representing Hispanic persons of any race.

Weaknesses:

- Data are provided using 5-year age groups only.
 - Aggregating individual years into 5-year age groups was important to achieving a reliable re-distribution across all ages (most notably for non-whites over age 45).
- Same weaknesses as the annual post-censal estimates because the data inputs are identical:
 - Does not provide any estimates for geographies lower than MCD (Town), such as tracts.
 - Demographic breakdowns are not available at the MCD (Town) level.
 - Demographic breakdowns that are available for County-level and up are not as complete as the decennial census.
 - Inputs (births, deaths, migration) are often delayed limiting true-to-date estimates
- Collapsing the Asian race group with the Native Hawaiian and Other Pacific Islander race group may be a weakness for some users.
- Collapsing all Hispanic persons into a single race*ethnicity category may be a weakness for some users.

Town ASRH:

- NCHS does not publish ASRH data at the town-level because the USCB does not publish data at the town-level.
- **CT DPH developed town-ASRH estimates for July 1, 2000 and July 1, 2010 only.**

**July 1, 2000 & July 1, 2010
town-level bridged-ASRH post-censal dataset available @
<http://www.ct.gov/dph/populationdata>:**

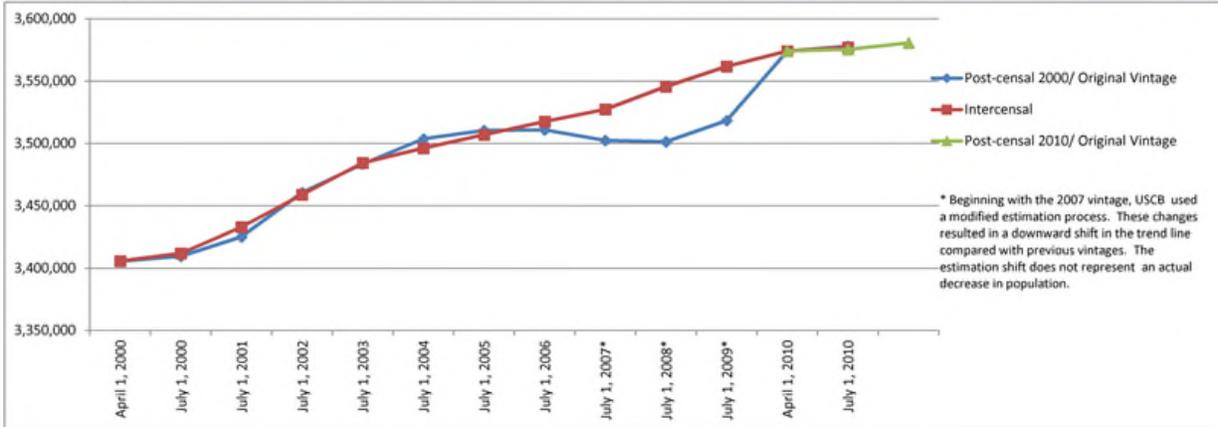
[Town: Population by Age, Sex, Race and Hispanic Ethnicity \(2000 & 2010\)](#)

Intercensal Estimates

As previously discussed, the USCB publishes annual, post-censal estimates of the population that use the most recent decennial census as the base for the estimation process. When another decennial census is completed, the USCB produces *intercensal* estimates. Intercensal estimates are created for the years between two decennial censuses when both the beginning (e.g., Census 2000) and ending (e.g., Census 2010) populations are known – hence the term *INTER-censal* versus the term *POST-censal*. Intercensals are produced once a decade by adjusting the existing time series of post-censal estimates for the entire decade to create a smooth transition from one decennial census count to the next. They differ from the post-censal estimates that are released annually because they rely on a mathematical formula that redistributes the difference between the April 1 post-censal *estimate* and April 1 census *count* for the end of the decade across the estimates for that decade.

Population Estimates: Comparison of Vintages and Intercensals

| Connecticut's Population Estimates | April 1, 2000 | July 1, 2000 | July 1, 2001 | July 1, 2002 | July 1, 2003 | July 1, 2004 | July 1, 2005 | July 1, 2006 | July 1, 2007* | July 1, 2008* | July 1, 2009* | April 1, 2010 | July 1, 2010 | July 1, 2011 |
|------------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|---------------|---------------|---------------|--------------|--------------|
| Post-censal 2000/ Original Vintage | 3,405,565 | 3,409,549 | 3,425,074 | 3,460,503 | 3,483,390 | 3,503,604 | 3,510,297 | 3,510,787 | 3,502,309 | 3,501,252 | 3,518,288 | 3,574,097 | 3,577,845 | |
| Annual Change | | | 0.46% | 1.03% | 0.66% | 0.58% | 0.19% | 0.01% | -0.24% | -0.03% | 0.49% | | 1.67% | |
| Intercensal | 3,405,650 | 3,411,777 | 3,432,835 | 3,458,749 | 3,484,336 | 3,496,094 | 3,506,956 | 3,517,460 | 3,527,270 | 3,545,579 | 3,561,807 | 3,574,097 | 3,577,073 | |
| Annual Change | | | 0.62% | 0.75% | 0.74% | 0.34% | 0.31% | 0.30% | 0.28% | 0.52% | 0.46% | | 0.43% | |
| Post-censal 2010/ Original Vintage | | | | | | | | | | | | 3,574,097 | 3,575,498 | 3,580,709 |
| Annual Change | | | | | | | | | | | | | | 0.15% |



The U.S. Census Bureau’s (USCB) intercensal estimates are published a few years after the latest decennial census. The USCB’s intercensal estimate series mirrors the post-censal estimates series. Likewise, the National Center for Health Statistics (NCHS) also publishes intercensal estimates that mirror their annual post-censal bridged-race ASRH estimates. This means that intercensal estimate series for 2000-2009 are available for the following post-censal datasets:

- Total population for State, County, and Town as of July 1 from USCB
- Demographic (ASRH) population estimates for State and County as of July 1 from NCHS

For dates when both post-censal and intercensal estimates are available, intercensal estimates are usually preferred.

Strengths:

- Uses decennial census figures for both the starting and ending points of the regression line so these are considered to better represent overall change throughout the decade.
- Only one intercensal series is provided
 - Intercensals are produced once, so they are not subject to the same vintaging that occurs with annual estimates
 - Revisions are possible. The USCB may publish a revised version of the intercensal dataset at a subsequent date.
- Maintains the same strengths as the dataset (annual, post-censal or annual bridged-ASRH) upon which the intercensal estimates are based.
- Intercensal series are provided for each of the USCB PEP's annual population estimate products.

Weaknesses:

- Smoothing the years between the two end points is a mathematical adjustment that may not accurately reflect the true population change from year to year.
- Maintains the same weaknesses as the dataset (annual, post-censal or annual bridged-ASRH) upon which the intercensal estimates are based.

Town ASRH:

- **The USCB publishes town-level intercensal population totals for July 1st annually; however, no demographic data is provided.**

Intercensal datasets available @ <http://www.ct.gov/dph/populationdata>:

Connecticut intercensal datasets are *not yet* available on the DPH website. Please contact Karyn Backus (karyn.backus@ct.gov) for access to intercensal datasets.

American Community Survey (ACS) Estimates

The American Community Survey (ACS) is an ongoing statistical survey by the U.S. Census Bureau that gathers information previously contained only in the long form of the decennial census. It is the largest survey other than the decennial census that the Census Bureau administers.

The ACS provides yearly estimates of the population and various demographic, economic, and health measures for all states, as well as all cities, counties, metropolitan areas, and population groups of 65,000 people or more. For smaller areas, it is necessary to combine multiple survey years to obtain reliable estimates: three survey years in areas with 20,000 to 65,000 people, and five survey years in areas with fewer than 20,000 people.

The quality of these samples was originally intended to match that of the decennial census long form, but because the sample size of the ACS is smaller than originally expected, ACS estimates are less precise than the comparable estimates from Census 2000 and prior decennial census years. Because these estimates are from survey data, Margins of Error (MOEs) are now at play. ACS provides MOEs for each estimate as well as directions for using MOEs.

Strengths:

- The only dataset that provides demographic data *below the county-level* on an ongoing basis.
 - Town, tract, block group, and block level data are available in 5-year datasets.
- The ongoing collection of sample data makes this survey the most current available.
 - Decennial census information becomes increasingly out-of-date as the decade progresses whereas ACS data collection is ongoing throughout the decade.
- Provides demographic estimates as of July 1st and is consistent with geographic *totals* for state, county, and town that are provided in the annual, post-censal estimates
 - The ACS constrains its population totals (but not demographics) with the PEP's population totals for the US, state, county and town geographies. This ensures that the two data sources are reporting the same overall population figures.

Weaknesses:

- Only the largest geographies (state, county, large cities) are available annually.
 - Town and tract data are limited to rolling 5-year estimates.
 - ACS provides the population as a single year representation of the 3-year or the 5-year period; when using a multiple year numerator (e.g., 2009-2011), ACS figures need to be multiplied by the same number of years (e.g., 3).
- The distributions of demographic characteristics within a given geography are NOT constrained. Although totals by geography are constrained, the age, sex, race, or ethnicity distribution in the ACS may differ from the annual, post-censal estimates.
- Demographic breakdowns are limited in comparison to other products.
 - Age is only available 5-year age groups and selected age categories. The older 5-year age groups are combined into 10-year age groups when crossed with race/ethnicity.
 - Race and ethnicity tables are provided independently. A race*ethnicity cross-tabulation is available in Table DP05 only.
 - When cross-tabulating race*ethnicity with age or sex, the only combined race/ethnicity group is White-NH. There is no estimate for non-Hispanic for any race group other than White when including age or sex in the breakdown.
 - These categories can make it difficult to use ACS data as denominators when attempting to analyze age*race independent of Hispanic ethnicity.
- Uses the decennial census race categories

- The Some Other Race category is collected but not reallocated
- The Multiple Race category is collected, but bridged-ASRH datasets are not available.
- Requires using Margins of Error (MOEs) when performing statistics
 - ACS provides documentation on calculating MOEs
 - When combining variables (e.g., "married" with "college degree"), MOEs need to be recalculated appropriately.

Town ASRH:

- ACS does not publish complete ASRH data at the town-level. Limited ASRH data is available for all CT towns for 5-year periods.
- **CT DPH discourages using the ACS data for town-level estimates that include demographics due to the large margins of error and the limited demographic breakdowns. We have found the margins of error are often larger than the estimate itself and can be erratic from year to year.**

Vintages

One of the first major hurdles of understanding the universe of population estimates is mastering the concept of *vintages*.

Wine Vintage:

1. The yield of wine or grapes ... during one season
2. Wine ... identified as to year and vineyard of origin
3. The year or place in which a wine was bottled

Population Vintage:

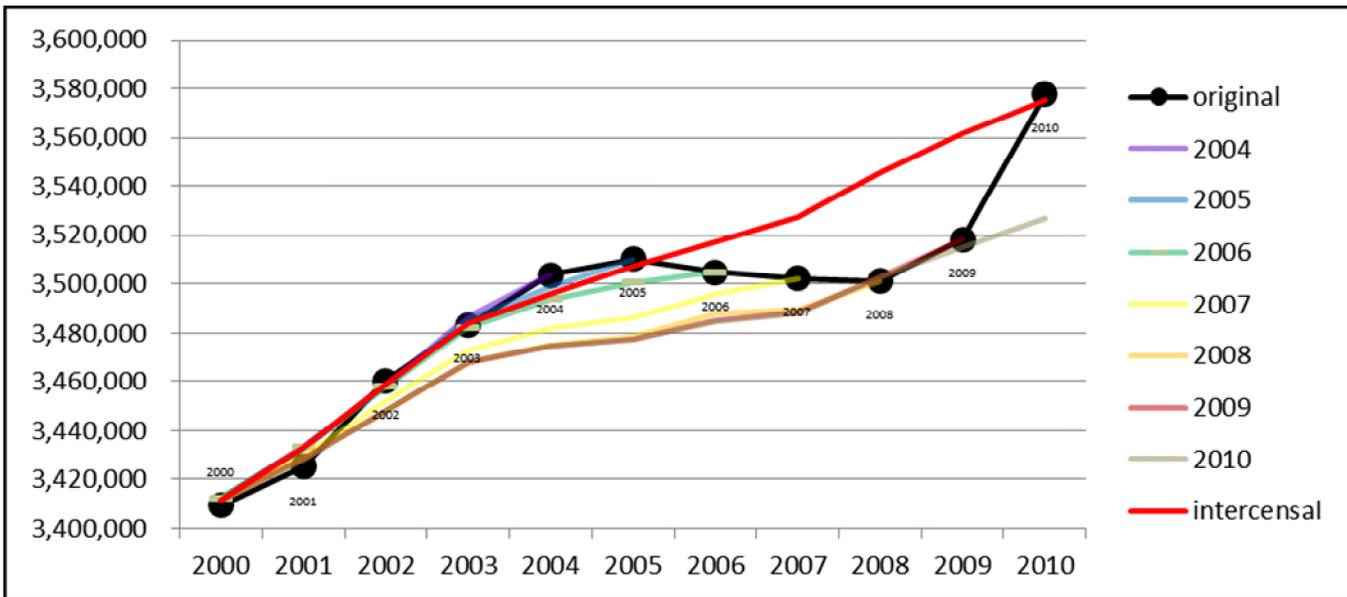
1. The estimated population using the inputs and methods for a given year
2. Population identified as to year and source of estimation
3. The year for which a population estimate was published

The *vintage year* (e.g., V2011) refers to the final year of the time series. With each new issue of July 1 estimates, the USCB revises estimates for all years back to the last census based on updated inputs, legal boundary changes, and changes to the estimate methodology. Since each year the USCB publishes revised estimates, each publication is noted by its vintage.

Using the table below, the population estimate for 2003 vintage 2008 (aka 2003v08) is 3,467,932. Notice that with each vintage, the population estimates for the previous year's change.

| Vintage of Estimate | Population Estimate Year | | | | | | | | | | |
|---------------------|--------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| original | 3,409,549 | 3,425,074 | 3,460,503 | 3,483,390 | 3,503,604 | 3,510,297 | 3,504,809 | 3,502,309 | 3,501,252 | 3,518,288 | 3,577,845 |
| 2004 | 3,412,262 | 3,433,243 | 3,459,006 | 3,486,960 | 3,503,604 | | | | | | |
| 2005 | 3,412,263 | 3,432,463 | 3,458,382 | 3,485,881 | 3,498,966 | 3,510,297 | | | | | |
| 2006 | 3,412,539 | 3,433,201 | 3,457,927 | 3,482,326 | 3,493,893 | 3,500,701 | 3,504,809 | | | | |
| 2007 | 3,411,990 | 3,429,770 | 3,451,867 | 3,472,964 | 3,481,890 | 3,486,490 | 3,495,753 | 3,502,309 | | | |
| 2008 | 3,411,714 | 3,428,208 | 3,448,261 | 3,467,932 | 3,475,351 | 3,478,714 | 3,487,896 | 3,489,868 | 3,501,252 | | |
| 2009 | 3,411,726 | 3,428,433 | 3,448,382 | 3,467,673 | 3,474,610 | 3,477,416 | 3,485,162 | 3,488,633 | 3,502,932 | 3,518,288 | |
| 2010 | 3,411,506 | 3,428,043 | 3,448,145 | 3,468,319 | 3,474,379 | 3,477,185 | 3,484,531 | 3,488,084 | 3,502,664 | 3,514,826 | 3,526,937 |
| intercensal | 3,411,777 | 3,432,835 | 3,458,749 | 3,484,336 | 3,496,094 | 3,506,956 | 3,517,460 | 3,527,270 | 3,545,579 | 3,561,807 | 3,575,498 |

The population estimates change with each vintage because of updated inputs, legal boundary changes, and changes to the estimate methodology. Included in the updated inputs is a new year of birth, death, and migration data. Adding a new point (i.e., a new year) to the series regression line causes the slope of the regression line to change which in turn moves the point estimates for each of the existing years. As you can see in the figure below, the colored lines shift as new points are added. As discussed in the annual, post-censal estimates section, the previously published population estimates are not wrong per se, the inputs for that vintage are just not as current as later vintages.



Changes in methodology can also cause noticeable changes in the slope of the estimate series. Looking at the figure above, the black line represents the original vintage estimate for each year from 2000-2010. In 2007, the USCB changed their estimation methodology. This change resulted in a more conservative estimate of population growth. Looking at the black line above, it appears that the 2007 population for CT was less than the 2006 population estimate. Did the CT population shrink in size from 2006 to 2007 rather than grow? The answer is NO. The new methodology is more conservative and when looking at the 2007 series line (in yellow), we see that the CT population did grow from 2006 to 2007. The change in methodology with the vintage 2007 series was applied to the entire regression line which results in a flatter slope from 2000 to 2007. Because each vintage is produced using inputs and methodologies specific to that vintage, it is important that each population estimate that you use is properly cited with the year and vintage of the estimate used.

From 2000 through 2009, a new vintage is produced each new year. Once the next decennial census is completed, the need for a “vintage” ceases. The regression line starts with decennial census 2000 and ends with decennial census 2010. There will be no new points added and updates to inputs for 2000-2009 are no longer applied. Essentially, the decade is closed to new information. At this time, with a census count for 2000 and a census count for 2010 and final information for the population change factors (births, deaths, migration) for 2000-2010, the USCB is able to compute a static *intercensal* time series for 2000-2010. The red line above represents the intercensal series.

Now that you understand vintages, you will see them crop up in other ways that are important. One of these important concepts is recognizing that the USCB constrains geographic totals across their data products for the same vintage. As seen in the figure below, all vintage 2011 *single-year* state estimates are 3,580,709 – even the ACS

survey estimate. In contrast, the totals by demographic subsets are not constrained. This means that the state, county, and town total populations will match across all datasets for each year (same vintage) but that the age, sex, race, and ethnicity makeup of each year may vary by data source.

The *total* population estimates by *geography* are constrained (see below that all vintage 2011 single-year state estimates are 3,580,709); however, the totals by demographic subsets are not constrained. This means that the state, county, and town *total* populations will match for each year (same vintage) but that the age, sex, race, and ethnic makeup of each year may vary by data source.

U.S. Census Bureau FactFinder Annual July 1, 2011 (vintage 2011)

| Geography | April 1, 2010 | Population Estimates (as of July 1) |
|-------------|------------------|---|
| Connecticut | Census 3,574,967 | Estimates Base 2010 3,575,498 2011 3,580,709 |

NCHS Bridged July 1, 2011 (vintage 2011)

| Age Group | White | Black | Hispanic | Asian | Other | Total |
|-----------|---------|---------|----------|---------|---------|-----------|
| 0-4 yrs | 10,880 | 12,205 | 2,502 | 2,427 | 48 | 28,062 |
| 5-9 yrs | 10,904 | 12,177 | 10,368 | 9,960 | 262 | 29,611 |
| 10-14 yrs | 10,919 | 12,158 | 12,789 | 12,379 | 274 | 36,519 |
| 15-19 yrs | 10,922 | 12,158 | 14,807 | 14,026 | 289 | 36,142 |
| 20-24 yrs | 10,925 | 12,158 | 16,816 | 15,839 | 292 | 35,839 |
| 25-29 yrs | 10,928 | 12,161 | 18,825 | 17,848 | 295 | 35,837 |
| 30-34 yrs | 10,931 | 12,164 | 20,834 | 19,857 | 298 | 35,840 |
| 35-39 yrs | 10,934 | 12,167 | 22,843 | 21,880 | 301 | 35,842 |
| 40-44 yrs | 10,937 | 12,170 | 24,852 | 23,917 | 304 | 35,845 |
| 45-49 yrs | 10,940 | 12,173 | 26,861 | 25,990 | 307 | 35,848 |
| 50-54 yrs | 10,943 | 12,176 | 28,870 | 27,163 | 310 | 35,851 |
| 55-59 yrs | 10,946 | 12,179 | 30,879 | 29,336 | 313 | 35,854 |
| 60-64 yrs | 10,949 | 12,182 | 32,888 | 31,509 | 316 | 35,857 |
| 65-69 yrs | 10,952 | 12,185 | 34,897 | 33,682 | 319 | 35,860 |
| 70-74 yrs | 10,955 | 12,188 | 36,906 | 35,855 | 322 | 35,863 |
| 75-79 yrs | 10,958 | 12,191 | 38,915 | 38,028 | 325 | 35,866 |
| 80-84 yrs | 10,961 | 12,194 | 40,924 | 40,201 | 328 | 35,869 |
| 85-89 yrs | 10,964 | 12,197 | 42,933 | 42,374 | 331 | 35,872 |
| 90-94 yrs | 10,967 | 12,200 | 44,942 | 44,547 | 334 | 35,875 |
| 95+ yrs | 10,970 | 12,203 | 46,951 | 46,720 | 337 | 35,878 |
| Total | 358,070 | 358,070 | 358,070 | 358,070 | 358,070 | 3,580,709 |

ACS DEMOGRAPHIC AND HOUSING ESTIMATES 2011 American Community Survey 1-Year Estimates

| Subject | Estimate | Margin of Error | Percent | Percent Margin of Error |
|------------------|-----------|-----------------|---------------|-------------------------|
| SEX AND AGE | | | | |
| Total population | 3,580,709 | ±1,391 | 3.580,709 (X) | ±0.1 |
| Male | 1,743,730 | ±1,391 | 51.2% | ±0.1 |
| Female | 1,836,979 | ±1,391 | 51.2% | ±0.1 |
| Under 5 years | 196,258 | ±1,187 | 5.5% | ±0.1 |
| 5 to 9 years | 221,622 | ±1,315 | 6.2% | ±0.2 |
| 10 to 14 years | 296,763 | ±1,496 | 8.3% | ±0.2 |

2011 ACS 1-year Estimates*

* Only the ACS 1-year estimates will be consistent because 3- and 5-year estimates are averages.

Artifact Due to Vintages

As discussed above, changes in the estimation methodology can create noticeable differences in the population figures for a particular year or perhaps for the decade as a whole. For this decade, these differences are most evident for 2007, 2008, and 2009 where the post-censal estimates are notably lower than intercensal estimates.

There may be times when a new rate is different that a previously published rate for the same time period. It is valuable for you to take pause and consider whether differences in the population estimates used for the denominator may be contributing to this rate change. Because the 2007-2009 post-censal data is lower than the intercensal figures for the same years, the calculated crude mortality rate (CMR) for Connecticut was lower when using the intercensal data. Had I been comparing the 2009 CMR (post-censal 2009v09) to the 2010 CMR (intercensal 2010), I would conclude that the CMR was lower in 2010 than it was in 2009. When using the same series for the denominator for

the 2009 CMR and the 2010 CMR, the 2010 CMR is slightly higher than the 2009 CMR. In most cases, the population variation from series to series is low and will have an insignificant effect on your analyses. A possible exception to this is the jump in the population change between vintage 2009 series and vintage 2010 series.

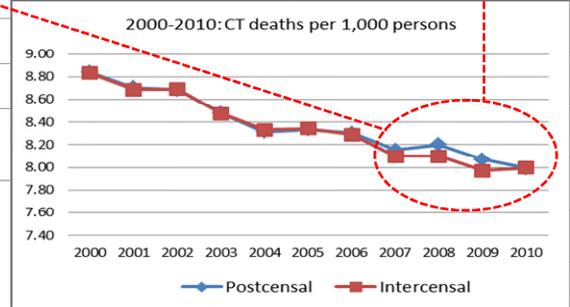
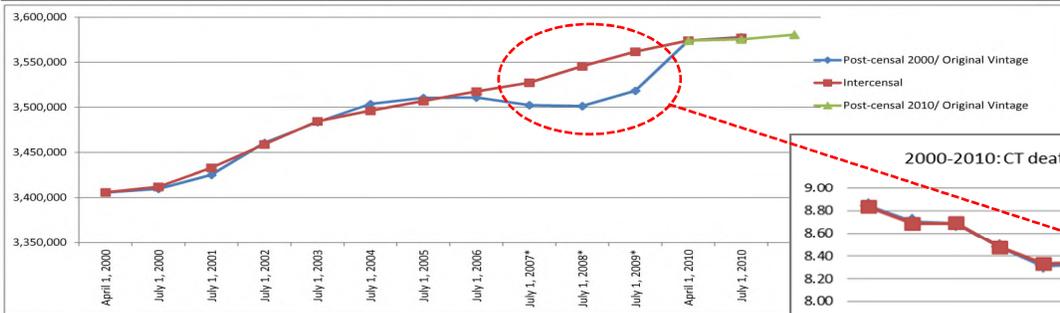
When comparing rates over time, verify the denominators used for the rates at each time point before making conclusions about the significance of the change

- As discussed in the vintages slide, the estimated population for a given year changes over time as inputs and methodologies are revised. Care should be taken to ensure that rate differences are not due to differing population denominators.
- For example, when evaluating the past decade for rate changes, the denominator for the rates should come from the same vintage/series. This prevents artifactual differences in population numbers (from varying inputs and methods) from impacting the results.
- Recognize that historical data and rates will use population estimates that differ from current data for the same year. Use caution when comparing your analyses to historical rates.
- Review and verify that all population estimates used in your analyses are appropriate for your project.

Population Estimates: Comparison of Vintages and Intercensals

| Connecticut's Population Estimates | April 1, 2000 | July 1, 2000 | July 1, 2001 | July 1, 2002 | July 1, 2003 | July 1, 2004 | July 1, 2005 | July 1, 2006 | July 1, 2007* | July 1, 2008* | July 1, 2009* | April 1, 2010 | July 1, 2010 | July 1, 2011 |
|------------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|---------------|---------------|---------------|--------------|--------------|
| Post-censal 2000/ Original Vintage | 3,405,565 | 3,409,549 | 3,425,074 | 3,460,503 | 3,483,390 | 3,503,604 | 3,510,297 | 3,510,787 | 3,502,309 | 3,501,252 | 3,518,288 | 3,574,097 | 3,577,845 | |
| Annual Change | | | 0.46% | 1.03% | 0.66% | 0.58% | 0.19% | 0.01% | -0.24% | -0.03% | 0.49% | | 1.67% | |
| Intercensal | 3,405,650 | 3,411,777 | 3,432,835 | 3,458,749 | 3,484,336 | 3,496,094 | 3,506,956 | 3,517,460 | 3,527,270 | 3,545,579 | 3,561,807 | 3,574,097 | 3,577,073 | |
| Annual Change | | | 0.62% | 0.75% | 0.74% | 0.34% | 0.31% | 0.30% | 0.28% | 0.52% | 0.46% | | 0.43% | |
| Post-censal 2010/ Original Vintage | | | | | | | | | | | | 3,574,097 | 3,575,498 | 3,580,709 |
| Annual Change | | | | | | | | | | | | | | 0.15% |

Rate change is an artifact of the denominator.



Citing Population Data

Using the citation “US Census Bureau” is not acceptable. Although each of the datasets discussed here originate with the U.S. Census Bureau, each dataset has a unique citation that properly documents the production of the dataset. The citation includes details about the data source that will allow the reader of your report to recreate the same analyses. Elements such as the vintage and date of publication are significant components when properly identifying the datasets. Do not skip the details in favor of a generic citation.

It is also important that you provide the citation that specifies where YOU accessed the data. For example, if you use the state population data that was published by CT DPH, you will use the CT DPH citation for that dataset even though the CT DPH cites the USCB as the data source.

The good news is that citations are usually provided. Each of the DPH population datasets is published with a recommended citation. Both the Census Bureau and NCHS provide recommended citations in the documentation associated with the dataset. The documentation files are located on the same webpages as the datasets. Data downloaded from American FactFinder can be difficult to cite because of the nature of the query system. The USCB has provided details on how to properly cite data from AFF:

<https://ask.census.gov/faq.php?id=5000&faqId=519>.

A recommended citation contains all of the information necessary to properly identify the dataset. The “recommended” qualifier applies to the format of the citation, not the information provided within the citation. For DPH, we use the American Psychological Association’s editorial style that many of the social and behavioral sciences have adopted to present written material in the field. Detailed information about the APA Style and how to format your citations in APA Style can be found at www.apastyle.org.

SAS Datasets

Selected population datasets are available in SAS format. If the SAS version of the dataset is not posted on the webpage near the excel version, you may contact Karyn Backus (karyn.backus@ct.gov) to see if one is available.

Population Statistics: Q & A

I need the CT population data for the state, a county, or a town.

Population tables are published on the DPH website:

[DPH HOME PAGE > STATISTICS & RESEARCH > POPULATION STATISTICS](#)

The population tables provided on-line are in a tabular format. If you would like the population data in non-tabular form (SAS, DBF, Excel) for analytic purposes, you may contact Karyn Backus directly to request a dataset.

The screenshot shows the website for the Connecticut Department of Public Health (DPH). The header includes the DPH logo with the tagline "Keeping Connecticut Healthy" and the text "DEPARTMENT OF PUBLIC HEALTH". A navigation bar contains links for Home, About Us, Publications, Forms, and Contact Us. On the left side, there is a profile for Dr. Jewel Mullen, Commissioner, and a sidebar menu with categories like "Services & Programs", "Regulation & Licensure", "Vital Records", "Statistics & Research", "News Room", and "Topics A - Z". Below the menu are several buttons: "TRINConnecticut", "H.I.P.A.A NOTICE", "HealthCare Reform", and "Check the Calendar". The main content area is titled "Population Statistics" and contains several sections: "Population Statistics Overview" (with a link to a detailed overview), "Annual July 1 Estimates" (explaining that these estimates are based on the calendar year midpoint), "Intercensal Estimates - coming soon" (explaining they are produced between decennial censuses), "Decennial Census Population Counts" (explaining the decennial census provides a true count), and "American Community Survey Estimates" (explaining the ACS provides demographic, social, and economic data).

DPH Population Statistics Webpage

The Population Statistics Overview puts details about each population data source in a single document. This is helpful for understanding the geographic and ASRH limitations of the different datasets.

The Overview also has a Q&A section to help the data user determine the best data source for their project.

Population Estimates - The calculated number of people living in an area as of a specified point in time, usually July 1st. The estimated population is calculated using a component of change model that incorporates information on natural increase (births, deaths) and net migration (net domestic migration, net international migration) that has occurred in an area since the latest decennial census.

Base Population - The population count or estimate used as the starting point in the estimates process. It can be the most recent updated Census count or the estimate for a previous date within the same vintage. The April 1, 2010 estimates base population may differ from the April 1, 2010 Census count due to legal boundary updates, other geographic program changes, and Count Question Resolution actions.

Postcensal Estimates - Population estimates produced for the years after a decennial census when only the beginning population is known. They are produced and revised each year. For dates when both postcensal and intercensal estimates are available, intercensal estimates are preferred.

Intercensal Estimates - Population estimates produced for the years between two decennial censuses when both the beginning and ending populations are known. They are produced once a decade by adjusting the existing time series of postcensal estimates for the entire decade to smooth the transition from one decennial census count to the next. They differ from the postcensal estimates that are released annually because they rely on a mathematical formula that redistributes the difference between the April 1 postcensal estimate and April 1 census count for the end of the decade across the estimates for that decade. For dates when both postcensal and intercensal estimates are available, intercensal estimates are preferred.

What are the USCB's Minor Civil Divisions?

While many of states throughout the nation use county as the primary political/municipal entity at the local level, about 20 states use smaller geographies (towns, townships) as the primary local governmental unit. These legally defined county subdivisions are referred to as *minor civil divisions* (MCDs) by the USCB. The USCB produces estimates at the MCD level (in addition to higher levels) for these states. In CT, there are 169 official towns and all 169 are MCDs. Incorporated places are subordinate to the MCDs in which they are located.

Why do the official CT annual estimates for state, county, or town differ from the USCB annual mid-year estimates?

Official CT estimates and USCB figures may differ for a variety of reasons.

- a) In some years, CT needed to make adjustments to the Census' estimates which resulted in the official CT estimates being slightly different than the Census estimates. All such adjustments are discussed in the Methods of Estimation section of our town-level population estimate reports or in the Notes section of the ASRH estimates data file.
- b) Also, different estimates for the same year may be due to comparing different estimate *vintages*. CT's official estimates use the first vintage available for a given year and are not revised/updated for later *vintages*. With each new issue of July 1 estimates, the USCB revises estimates for all years back to the last census. Most administrative record data sources lag the current estimate year (by as much as two

years), therefore, the USCB projects the data for the current year based on past years' data. As updated data become available, they revise the projected input data so that each *vintage's* estimates are always based on the most recent data available. Revisions to estimates also incorporate changes in methodology and legal boundary changes. Since each year the USCB publishes revised estimates, each publication is noted by its *vintage*.

Why do the official CT annual estimates for state, county, or town differ from the ACS estimates?

The annual post-censal USCB estimates and the ACS estimates use different methods.

- a) The official CT annual estimates are based on the annual post-censal USCB estimates. The Population Estimates Program (PEP) at the Census Bureau develops and prepares estimates of the population by age, sex, race, and Hispanic origin for the nation, states, counties, and towns. These estimates are consistent with the decennial census residence definition of usual residence and represent the Census Bureau's official estimates of updated census counts for these areas. The USCB annual estimates are used for a variety of purposes including the annual allocation of over \$400 billion in federal funds, as denominators for statistical indicators, and the basis for program planning and development.
- b) Variants of these estimates are used as controls for the American Community Survey (ACS). The population levels by age, sex, race, and Hispanic origin for states and counties, the population totals for places, and the number of housing units for states and counties which result from the ACS will not agree with the USCB official estimates. This is because the weighting specifications employed by the ACS use various aggregations of the estimates for states and counties.

Thus, ACS estimates of population levels and housing unit estimates should not be used for state, county, or town when then USCB annual population estimates can be used instead.

Why do the ACS tract totals differ from the 2010 Census tract totals?

The 2010 Census tracts totals represent the population count as of April 1, 2010 for each tract. The ACS tract totals represent an estimate by ACS for each tract as of July 1st over a five year period that is averaged together.

I need tract level population counts for 2010.

The decennial census will provide population counts by census tract for April 1, 2010. USCB tract level estimates are not available for any other time point. ACS provides tract level estimates for July 1 for 5-year estimates only.

I need block group level economic data for 2010.

The decennial census will not provide economic data. Economic data is available through the ACS or other Census Bureau products. Block group level data will be available for 5-year estimates only.

I need town level data by age, sex, race, or ethnicity for each year 2010-2019.

Annual estimates with demographics are not available at the town level for individual years. Annual population estimates with demographics at the town level are not

published by the USCB. The ACS will publish town level estimates with demographics, but only the 5-year estimates will provide all 169 CT towns.

I want to find the population count for a particular demographic group at both the town level and tract level.

If you are going to compare town and tract for the same analysis, the same data source (census 2010 or ACS) should be used. Both data sources will provide demographic data. The Census 2010 data will be for April 1, 2010. The ACS data for town or tract will be for 5-year estimates only.

I want to compare the change in the rate of a health indicator for a particular geography over a series of years. Which population figures should I use for denominators?

Each year the USCB publishes a new series (*vintage*) of estimates in which the entire series of years are updated to reflect the most current input data and methodology. When using CT population estimates as denominators for rates (e.g., prevalence rates) for several years, it is important to use the same vintage for all years. For example, if you want to calculate the birth rate for each town by year for the decade, the same *vintage* of town population estimates should be used to pull the denominator counts for each year.

The USCB's annual population estimates for CT are available at several levels of geography, such as state, county, town/MCD, and incorporated place/borough. The USCB does NOT publish annual estimates based on decennial census geographies such as tracts, block groups, or blocks. If looking to calculate rates for small areas such as these, you will need to use ACS data. Be aware that the ACS estimate is independent of the USCB annual estimates and derived using different inputs and methodologies. If seeking to compare rates between different geography levels, we recommend that you stay within the same estimate source: if comparing town rates with tract rates, both denominators should come from ACS since only ACS provides tract-level data.